

VU Research Portal

Causes of reporting bias

van der Steen, Jenny T; Ter Riet, Gerben; van den Bogert, Cornelis A; Bouter, Lex M

published in
F1000Research
2019

DOI (link to publisher)
[10.12688/f1000research.18310.2](https://doi.org/10.12688/f1000research.18310.2)

document version
Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

van der Steen, J. T., Ter Riet, G., van den Bogert, C. A., & Bouter, L. M. (2019). Causes of reporting bias: a theoretical framework. *F1000Research*, 8, 280. [280]. <https://doi.org/10.12688/f1000research.18310.2>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl



RESEARCH NOTE

REVISED Causes of reporting bias: a theoretical framework [version 2; peer review: 2 approved]

Jenny T van der Steen ^{1,2}, Gerben ter Riet ^{3,4}, Cornelis A van den Bogert⁵,
Lex M Bouter ^{6,7}

¹Department of Public Health and Primary Care, Leiden University Medical Center, Hippocratespad 21, Gebouw 3, Leiden, 2300 RC Leiden, The Netherlands

²Department of Primary and Community Care, Radboud university medical center, Geert Grooteplein Noord 21, 6500 HB Nijmegen, The Netherlands

³ACHIEVE Centre for Applied Research, Amsterdam University of Applied Sciences, Tafelbergweg 51, Amsterdam, 1105 BD Amsterdam, The Netherlands

⁴Department of Cardiology, Amsterdam University Medical Center (location Meibergdreef), University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

⁵Apotheek Boekel, Kerkstraat 35, Boekel, 5427 BB, The Netherlands

⁶Department of Epidemiology and Biostatistics, Amsterdam University Medical Centers, location VUmc, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands

⁷Department of Philosophy, Faculty of Humanities, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

v2 First published: 12 Mar 2019, 8:280 (
<https://doi.org/10.12688/f1000research.18310.1>)

Latest published: 17 Jul 2019, 8:280 (
<https://doi.org/10.12688/f1000research.18310.2>)

Abstract

Reporting of research findings is often selective. This threatens the validity of the published body of knowledge if the decision to report depends on the nature of the results. The evidence derived from studies on causes and mechanisms underlying selective reporting may help to avoid or reduce reporting bias. Such research should be guided by a theoretical framework of possible causal pathways that lead to reporting bias. We build upon a classification of determinants of selective reporting that we recently developed in a systematic review of the topic. The resulting theoretical framework features four clusters of causes. There are two clusters of necessary causes: (A) motivations (e.g. a preference for particular findings) and (B) means (e.g. a flexible study design). These two combined represent a sufficient cause for reporting bias to occur. The framework also features two clusters of component causes: (C) conflicts and balancing of interests referring to the individual or the team, and (D) pressures from science and society. The component causes may modify the effect of the necessary causes or may lead to reporting bias mediated through the necessary causes. Our theoretical framework is meant to inspire further research and to create awareness among researchers and end-users of research about reporting bias and its causes.

Keywords

Causality, publication bias, questionable research practice, reporting bias, research design, selective reporting

Open Peer Review

Reviewer Status

	Invited Reviewers	
	1	2
REVISED		
version 2	report	report
published 17 Jul 2019		
version 1		
published 12 Mar 2019	report	report

1 **Ksenija Bazdaric** , University of Rijeka, Rijeka, Croatia

2 **Arnaud Vaganay** , Meta-Lab, London, UK
National Centre for Social Research (NatCen), London, UK

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the **Science Policy**
Research gateway.

Corresponding author: Jenny T van der Steen (jtvandersteen@lumc.nl)

Author roles: **van der Steen JT:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **ter Riet G:** Conceptualization, Formal Analysis, Methodology, Supervision, Validation, Writing – Review & Editing; **van den Bogert CA:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Validation, Visualization, Writing – Review & Editing; **Bouter LM:** Conceptualization, Formal Analysis, Methodology, Supervision, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The writing of the article was supported by personal grants to JTvdS from the Netherlands Organisation for Scientific Research (NWO; Innovational Research Incentives Scheme: Vidi grant number 917.11.339), and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Consolidator grant agreement No 771483), and by Leiden University Medical Center, Leiden, The Netherlands.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 van der Steen JT *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution Licence**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: van der Steen JT, ter Riet G, van den Bogert CA and Bouter LM. **Causes of reporting bias: a theoretical framework [version 2; peer review: 2 approved]** F1000Research 2019, 8:280 (<https://doi.org/10.12688/f1000research.18310.2>)

First published: 12 Mar 2019, 8:280 (<https://doi.org/10.12688/f1000research.18310.1>)

REVISED Amendments from Version 1

In the new version, we provided more background on how the theoretical framework on causes of reporting bias was developed. We emphasized that the 12 categories of determinants were derived via a qualitative inductive content analysis of the literature. We illustrated how the literature also inspired to theorize possible relationships between the determinant categories and to distinguish four clusters of determinants. We clarified that we used existing epidemiologic terminology to label relationships between the clusters only afterwards.

The framework presents a generic model in which behavior of individuals is prominent in the two necessary causes 'Motivations' and 'Means.' It also shows the impact of societal factors on the outcome reporting bias which cannot be brought about by behavior of a single person. We added an example of how editors and authors are both actors in the determinant categories of 'Academic publication system hurdles' and the extent to which authors have 'Doubts about reporting being worth the effort.' We also added the example of replication studies to show how promoting these studies could affect reporting bias via all four clusters of determinant categories.

We recognize that the article may have neglected designs other than experimental designs, and we added that the model could also be useful to help assess possible confounding factors in observational research on reporting bias. We also recognize that the evidence base for causes of reporting bias is limited, calling for studies to increase the evidence base considering relevant determinants, and refinement of theory for particular fields.

See referee reports

Background

The problem of selective reporting and research on reporting bias

Selective reporting of research findings presents a large-scale problem in science, substantially affecting the validity of the published body of knowledge (Bouter *et al.*, 2016; Dwan *et al.*, 2014; van den Bogert *et al.*, 2017). Reporting bias (publication bias or outcome reporting bias) occurs when the decision to report depends on the direction or magnitude of the findings. In clinical research, registration of trials prior to data collection is used to prevent selective reporting (Chan *et al.*, 2017; Gopal *et al.*, 2018). However, it is insufficiently effective because despite registration or publication of the study protocol, trial results often remain partially or completely unpublished (Jones *et al.*, 2013) and selective reporting of "positive findings" also occurs among trials registered at, for example, clinicaltrials.gov (Dechartres *et al.*, 2016).

Although many epidemiological studies have described the occurrence or phenomenon of selective reporting, very few studies have targeted its causes. In particular there is little high-quality evidence on effective interventions. To develop effective interventions against reporting bias, we need a good understanding of possible contributions of actors involved (such as academic environment, editors, researchers) and of possible mechanisms. We also need clear hypotheses of how causes may be interrelated.

Basis for a theoretical causal framework: hypothesized determinants of selective reporting and their interrelationships

We recently developed a taxonomy of putative determinants of selective reporting based on themes abstracted from the literature (van der Steen *et al.*, 2018). We used an inductive approach of qualitative content analyses of empirical and non-empirical studies until we reached saturation, which indicates that the categories likely cover all important putative determinants of selective reporting. This resulted in 12 categories (Table 1).

In the literature review we also found some instances of hypothesized effect modification of the determinants of selective reporting, so that the effects of determinants are assumed not to be simply additive. For example, "Outcomes could be deemed post hoc to have little clinical relevance if they fail to show significant findings and may thus be omitted *when* accommodating space limitations" (Chan & Altman, 2005). In this case, a preference, namely statistically significant findings, combined with editorial practices lead to reporting bias. Similarly, Ioannidis (2005) hypothesized that a focus on preferred, positive findings could result in reporting of non-reproducible findings (only) if there is also an opportunity to do so through flexibility in study designs and freedom in reporting on it. That is, he concludes that "The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true" because "Flexibility *increases the potential* for transforming what would be 'negative' results into 'positive' results."

A framework of possible causal pathways to reporting bias

Motivations and means

Based on what we found in the literature and along the above lines, we hypothesize that the combination of two of the most common categories in our review (van der Steen *et al.*, 2018) — i.e., focusing on preferred findings and employing a poor or flexible study design, suffices to cause bias through selective reporting. Through multiple discussions in our team featuring experience in both qualitative and quantitative research, we inductively derived Figure 1 which shows how, as a next step, we identified and presented clusters covering these and the ten other categories of determinants and their possible interrelationships. We then added qualifications of the relationship inspired by Rothman's (1976) framework of necessary, sufficient and component causes. The two categories are part of clusters A (motivations) and B (means). We view both clusters A and B as necessary causes, that is, they are both part of any sufficient cause of reporting bias. This does not mean that reporting bias will always be the result of presence of A and B because effects can be mitigated by interventions and modified by component causes. Applying more epidemiological terms to the generic model we developed, there is also effect modification between A and B because reporting bias is not possible with A or B alone. Note that a preference for a particular outcome is not necessarily the authors' preference; it may also be that of a reviewer or editor. In addition to clusters A and B, we propose clusters C and D containing categories of component causes which are discussed in the next section.

Table 1. Twelve categories of determinants of selective reporting. (Modified from the taxonomy of determinants presented in Table 3 in: Determinants of selective reporting: A taxonomy based on content analysis of a random selection of the literature. van der Steen JT *et al.* PLoS One. 2018 Feb 5;13(2):e0188247. doi: [10.1371/journal.pone.0188247](https://doi.org/10.1371/journal.pone.0188247).)

Determinant category	Description	Examples
A. Motivations		
Preference for particular findings	A particular preference motivates a focus on finding results that match preferences, mostly statistically significant or otherwise positive findings, wishful thinking and acting	Significance chasing, finding significant results, larger effect size, suppressing publication of unfavourable results, not being intrigued by null findings
Prejudice (belief)	A conscious or unconscious belief that may be unfounded, and of which one may or may not be aware	Prior belief about efficacy of treatment, author reputation or gender bias in the phase of review
B. Means		
Opportunities through poor or flexible study design*	Attributes of study design relating to power and level of evidence provide much leeway in how studies are performed and in interpretation of their results	Not a controlled or blinded study, study protocol unavailable, small sample size
Limitations in reporting and editorial practices	Constraints and barriers to the practice of reporting relevant detail	Journal space restrictions, author writing skills
C. Conflicts and balancing of interests		
Relationship and collaboration issues	Intellectual conflict of interest between reporting and maintaining good relationships	Disagreements among co-authors and between authors and sponsors, sponsors prefer to work with investigators who share the sponsor's position
Dependence upon sponsors	Financial conflict of interest resulting in lack of academic freedom	Requirements and influence of funding source with financial interests in study results
Doubts about reporting being worth the effort	Weighing investment of time and means versus likelihood of gain through publication	Anticipating disappointment of yet another rejection or low chances of acceptance of a manuscript, belief that findings are not worth the trouble
Lack of resources, including time	Insufficient manpower or finances	Lack of time resulting from excessive workload, or lack of personnel due to life events
D. Pressures from science and society		
Academic publication system hurdles	Various hurdles to full reporting related to submission and processing of manuscripts (other than reporting) including those that represent an intellectual conflict of interest	Solicited manuscripts, authors indicating non-preferred reviewers, editor's rejection rate
High-risk area and its development	Area of research or discipline or specialty including its historical development and competitiveness, the currently dominant paradigms and designs, and career opportunities	Ideological biases in a research field, area with much epidemiological research versus clinical or laboratory research ("hard sciences"), humanities, experimental analytic methods, "hot" fields, publication pressure in the specific field
Unfavourable geographical or regulatory environment	Geographical or regulatory environment that affects how research is being performed	Continents under study included North America, Europe and Asia; few international collaborations; no governmental regulation of commercially sponsored research, ethics in publishing enterprise
Potential harm	Publishing data can harm individuals	Risk of bioterrorism, or confidentiality restriction

*With study design, we mean broader design issues than just type of research design, including also definitions, outcomes, analytic plans etc.

Poor or flexible study design may offer the means for selective reporting in addition to limitations in reporting and editorial practices (cluster B in Figure 1). In parallel, we placed "prejudice" in cluster A together with "preference for particular findings" because both may, whether consciously or not, represent a motivation for behaviour that leads to reporting bias. The possible motivations, wishes and beliefs in cluster A are different concepts that may result in "wishful

thinking" (Bastardi *et al.*, 2011) and in motivated reasoning around the interpretation of scientific findings (e.g. to serve political interests; Colombo *et al.*, 2016; Kraft *et al.*, 2015). Persons may or may not be fully aware of their motivations and the resulting behaviour may or may not be intentional (Greenland, 2009). Dickersin & Min (1993) stated that at the root of reporting bias may lay the very natural tendency to make public our successes. Success can be defined in different,

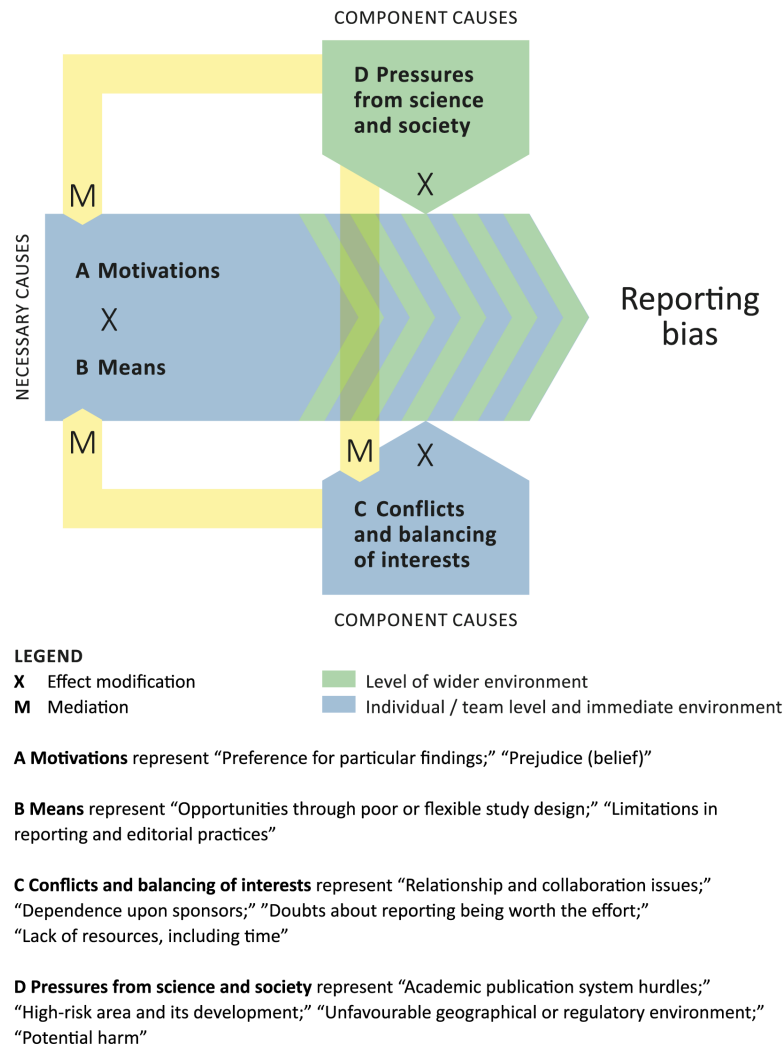


Figure 1. A theoretical framework for reporting bias. Bullet points indicate the 12 categories of determinants of selective reporting subsumed under four higher-level clusters A, B, C, and D. Note that the figure implies effect modification between A and B (necessary causes) because there will be no reporting bias with A or B alone. Effect modification ("X") may also occur by C or D and thus make the joint effect of A and B stronger. Mediation ("M") may occur if the necessary causes (A and B) mediate the effect of D. Mediation may also occur if C mediates the effects of D on A and B, which in its turn leads to reporting bias.

or even opposite ways as suggested by Rosenthal and Rubin cited by Preston *et al.* (2004) whose article was part of our review: "[E]arly in the history of a research domain results in either direction are important news but that later, when the preponderance of evidence has supported one direction, significant reversals are often more important news than further replications."

The pertinence of the second necessary cause (cluster B)—multiple opportunities to select what to analyse or report—is illustrated by the many degrees of freedom that researchers have but should not be tempted to use (in performing psychological research: Wicherts *et al.*, 2016). The necessary causes thus represent having a motive (preference or prejudice; cluster A) and the means (opportunities in study design or reporting; cluster B). Together they may form a sufficient cause for reporting bias.

Obviously, researchers and editors are key stakeholders because commonly they co-determine what will be reported. It can be argued that researchers are the most important because a single editor's decision is not decisive for non-publication or selective publication. Researchers are actors in three of the four categories in clusters A and B that represent the necessary causes, while editors are key players in only one category (in cluster B; Figure 1). Note that we assume actors in the field are capable of effective action.

Conflicts and balancing of interests and the wider environment

In the review, we found that after a series of rejections researchers may doubt whether reporting is worth the effort given lack of resources such as time. Balancing effort and output is placed in cluster C (component cause conflicts and balancing of interests; Figure 1). Cluster C also includes relationship

and collaboration issues and dependence upon sponsors. Cluster C thus represents conflicts of interests, individuals and teams juggling with harmony in relationships and time investments.

Other component causes represent pressures from the wider environment, such as from science and society (cluster D). The individual researcher has less control over type C, and in particular type D causes, than over motivations (A) and means (B). C and D cannot fully control or explain individuals' decisions, but they may shape motivations (A) and means (B). When this is the case the effect on reporting bias of the categories in cluster C or D is mediated through the categories contained in cluster A or B. For example, important news is selectively reported but what is deemed important news is shaped by the development within a scientific domain (cluster C; Preston *et al.*, 2004). Also, researchers' collaborations or relations with sponsors may nudge them to selectively report the preferences of others. A final example is academic publication system hurdles (cluster D) and dependence upon sponsors (cluster C) leading to reporting bias through their impact on the combination of a preference for positive findings and the opportunities that flexible designs offer.

Discussion

We propose a broad theoretical framework of reporting bias by relating and ordering 12 determinant categories that we derived from the literature (van der Steen *et al.*, 2018). We inductively combined these categories in four clusters (A–D) using existing epidemiologic terminology to label relationships.

The model is more refined than we anticipated when we wrote a protocol to develop a taxonomy of determinants of selective reporting and their interrelationships. We then expected a central role for preferences for particular “positive” findings only (van der Steen *et al.*, 2018 Supplement 1, Figure 1). However, having the means is necessary too. Although the determinants in our model are mostly based on research in the biomedical area, the model fits well with the “Desire-Belief-Opportunity” (DBO) model that analytical sociologists use to explain various phenomena (Hedström, 2005) and which we came across after having developed our theoretical framework. Desire and Belief concur with the two motivations in cluster A, while opportunities (alternative actions available to the actor) represent the means in cluster B.

Theory may guide the development of interventions as research often does not systematically consider contextual and individual factors that influence delivery of an intervention. Thus, theory may help avoid an ad hoc or data-driven approach to attempts to reduce reporting bias. It may also help explain some other phenomena, for example, problems with replicability which are partly caused by selective reporting. Replication studies can effect the four clusters A–D. They can impact on Motivations when e.g., researchers more often aim at study results that are likely replicated, or when researchers conducting replication studies are more open to, or working towards, null results. They can also impact on Means, e.g. the rise of

specific journals that support publishing replicated studies, and on Conflicts and balancing of interests (e.g. earmarked resources for replication studies becoming available), and Pressures from science and society (e.g. less creative and innovative but rigorous research becoming more salonfähig).

Although one might assume that interventions addressing reporting bias effectively will be complex, the removal of a single necessary cause is obviously effective. For example, a potentially very effective measure that funders and (medical) ethics committees could adopt is systematic monitoring of all written research outputs and comparing the outcomes reported therein to the corresponding research protocols and statistical analysis plans and potential amendments. This would require that these organizations make submission of such documents to them or to a publicly available repositories mandatory, in addition to requiring submission of a research protocol or study registration. For this, automated or manual comparing protocols to publications is needed (ter Riet & Bouter, 2016; Wright *et al.*, 2018). In the jargon of this paper, this approach would eliminate the necessary cause ‘Means.’ Given suitable negative reinforcements (punishments, ‘blacklist’) following incomplete reporting, such measures may also reduce motivation to report selectively. Similarly, elements from the component causes contained in cluster C and D that are highly prevalent and strongly modify the combined effect of cluster A and B may be prioritized targets. Mediators can also be good candidates for intervention. For example, component causes contained in cluster C may mediate the impact of elements of D on elements of clusters A or B. The model may also assist in assessing potential confounding factors in observational work in which associations between a specific determinants and reporting bias is assessed.

In addition to informing the development of interventions that are subsequently evaluated, our framework may also help to identify high risk scientific fields. For example, areas where designs offer considerable flexibility or where the researchers' degrees of freedom are combined with strong beliefs or a mission to disseminate particular outcomes (Ioannidis, 2005). The model also shows that for example editors may influence the outcome in multiple ways; first, directly via Means (the category Limitations in reporting and editorial practices). Second, editors as a collective affect the rejection rates through Academic publication system hurdles (the category of Pressures from science and society), but also the extent to which authors find efforts to publish worthwhile (category of cluster Conflicts and balancing of interests). The latter is illustrated by the personal account of Speyer (2018).

Currently, the evidence for the theoretical framework is limited. Based on research, our theoretical framework may need to be adapted. Motive and Means may be stable clusters but the C and D type causes may change as science changes. Future work may also help to refine the framework's relevance for specific disciplinary fields (e.g., non-clinical biomedical research). Further empirical research is needed to specify, for example, what could

be an optimal level of flexibility for a particular field and study design. Nevertheless, because the causal pathways seem plausible, were derived from the literature on selective reporting and is congruent with theory developed in the social sciences (Hedström, 2005), we feel that the current work can already help to design further research on the effectiveness of interventions.

Data availability

Underlying data

PLOS ONE Supplement 2 to article van der Steen *et al.*, 2018. Determinants of selective reporting abstracted from the selected literature. “S2 File. Dataset with determinants.” In Excel available from: <https://doi.org/10.1371/journal.pone.0188247.s003> (van der Steen *et al.*, 2018)

PLOS ONE Supplement 3 to article van der Steen *et al.*, 2018. Categories of determinants of selective reporting with literature references. “S3 file. References to the 64 articles included in the determinant analysis, per category.” In Word available

from: <https://doi.org/10.1371/journal.pone.0188247.s004> (van der Steen *et al.*, 2018)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Grant information

The writing of the article was supported by personal grants to JTVdS from the Netherlands Organisation for Scientific Research (NWO; Innovational Research Incentives Scheme: Vidi grant number 917.11.339), and from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Consolidator grant agreement No 771483), and by Leiden University Medical Center, Leiden, The Netherlands.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Bastardi A, Uhlmann EL, Ross L: **Wishful thinking: belief, desire, and the motivated evaluation of scientific evidence.** *Psychol Sci.* 2011; **22**(6): 731–2.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bouter LM, Tijdink J, Axelsen N, *et al.*: **Ranking major and minor research misbehaviors: Results from a survey among participants of four World Conferences on Research Integrity.** *Res Integr Peer Rev.* 2016; **1**: 17.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chan AW, Altman DG: **Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors.** *BMJ.* 2005; **330**(7494): 753.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chan AW, Pello A, Kitchen J, *et al.*: **Association of trial registration with reporting of primary outcomes in protocols and publications.** *JAMA.* 2017; **318**(17): 1709–1711.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Colombo M, Bucher L, Inbar Y: **Explanatory judgment, moral offense and value-free science.** *Rev Philos Psychol.* 2016; **7**(4): 743–63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dechartres A, Bond EG, Scheer J, *et al.*: **Reporting of statistically significant results at ClinicalTrials.gov for completed superiority randomized controlled trials.** *BMC Med.* 2016; **14**(1): 192.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dickersin K, Min YI: **NIH clinical trials and publication bias.** *Online J Curr Clin Trials.* 1993; Doc No 50.
[PubMed Abstract](#)
- Dwan K, Altman DG, Clarke M, *et al.*: **Evidence for the selective reporting of analyses and discrepancies in clinical trials: A systematic review of cohort studies of clinical trials.** *PLoS Med.* 2014; **11**(6): e1001666.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gopal AD, Wallach JD, Aminawung JA, *et al.*: **Adherence to the International Committee of Medical Journal Editors’ (ICMJE) prospective registration policy and implications for outcome integrity: a cross-sectional analysis of trials published in high-impact specialty society journals.** *Trials.* 2018; **19**(1): 448.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Greenland S: **Accounting for uncertainty about investigator bias: disclosure is informative.** *J Epidemiol Community Health.* 2009; **63**(8): 593–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hedström P: **Dissecting the Social - On the Principles of Analytical Sociology.** Cambridge Cambridge University Press. 2005.
[Reference Source](#)
- Ioannidis JP: **Why most published research findings are false.** *PLoS Med.* 2005; **2**(8): e124.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jones CW, Handler L, Crowell KE, *et al.*: **Non-publication of large randomized clinical trials: Cross sectional analysis.** *BMJ.* 2013; **347**: f6104.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kraft PW, Lodge M, Taber CS: **Why People “Don’t trust the evidence”: Motivated reasoning and scientific beliefs.** *Ann Am Acad Polit Soc Sci.* 2015; **658**(1): 121–33.
[Publisher Full Text](#)
- Preston C, Ashby D, Smyth R: **Adjusting for publication bias: Modelling the selection process.** *J Eval Clin Pract.* 2004; **10**(2): 313–22.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rothman KJ: **Causes.** *Am J Epidemiol.* 1976; **104**(6): 587–92.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Speyer H: **Discovering the value of a “failed” trial.** *European Science Editing.* 2018; **44**(4): 80–82.
[Publisher Full Text](#)
- ter Riet G, Bouter LM: **How to end selective reporting in animal research. In Animal models for human cancer: discovery and development of novel therapeutics.** Wiley: Weinheim. ISBN 9783527339976. 2016.
[Publisher Full Text](#)
- van den Bogert CA, Souverein PC, Brekelmans CT, *et al.*: **Primary endpoint discrepancies were found in one in ten clinical drug trials. Results of an inception cohort study.** *J Clin Epidemiol.* 2017; **89**: 199–208.
[PubMed Abstract](#) | [Publisher Full Text](#)
- van der Steen JT, van den Bogert CA, van Soest-Poortvliet MC, *et al.*: **Determinants of selective reporting: A taxonomy based on content analysis of a random selection of the literature.** *PLoS One.* 2018; **13**(2): e0188247.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wicherts JM, Veldkamp CL, Augusteijn HE, *et al.*: **Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking.** *Front Psychol.* 2016; **7**: 1832.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wright D, Williams E, Bryce C, *et al.*: **A novel approach to sharing all available information from funded health research: The NIHR Journals Library.** *Health Res Policy Syst.* 2018; **16**(1): 70.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:



Version 2

Reviewer Report 27 August 2019

<https://doi.org/10.5256/f1000research.21747.r51330>

© 2019 Vaganay A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Arnaud Vaganay 

¹ Meta-Lab, London, UK

² National Centre for Social Research (NatCen), London, UK

I thank the authors for their response to my comments and for addressing my concerns/suggestions.

I am now satisfied with the article and leave it to the reader to assess the framework for themselves.

I hope this article will trigger more research on this fascinating subject.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Meta-research; research integrity; systematic reviews; social sciences; science policy; education; experimental and quasi-experimental methods.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 14 August 2019

<https://doi.org/10.5256/f1000research.21747.r51331>

© 2019 Bazdaric K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ksenija Bazdaric 

Department of Medical Informatics, Faculty of Medicine, University of Rijeka, Rijeka, Croatia

I approve the manuscript. The model is a specific one, does not include some factors which I have mentioned in my first report and still has to be tested.

I am happy with the revisions you have done.
Good luck in your endeavours.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Research ethics, psychology, medical informatics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 20 May 2019

<https://doi.org/10.5256/f1000research.20029.r47229>

© 2019 Vaganay A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Arnaud Vaganay

¹ Meta-Lab, London, UK

² National Centre for Social Research (NatCen), London, UK

I was pleased to review this manuscript. I assessed the proposed theoretical framework based on the following criteria: (1) utility, (2) comprehensiveness, (3) parsimony, (4) testability, (5) heurism, and (6) scope. Ironically, these criteria are not part of an established theoretical framework; they only reflect a brief review of the literature on the subject.

I would rate the utility of the proposed framework as high. As far as I know, this is one of the first attempts to synthesize the literature on the factors driving reporting bias. Additionally, and as mentioned by the authors, understanding these factors is essential to the development of mitigating measures.

I would rate the heurism (i.e. evidence base) of the framework as low. Granted, there is now a large body of literature on the prevalence of reporting bias and on the possible factors driving it. However, virtually none of these studies are experimental, so all the relationships found between the occurrence of reporting bias and the 12 categories are at best correlational – not causal. This is, in my view, the most fundamental flaw of the framework. The good news is, it could be easily addressed by changing the title of the manuscript from “Causes of reporting bias” to “Drivers of reporting bias”.

Related to the above-mentioned point, I would rate the testability of the framework as low. The only testable driver of reporting bias is “financial conflict of interests” (FCOI). Most other drivers (prejudice, relationship and collaboration issues, doubts about reporting being worth the effort, etc.) would be hard to test empirically. Virtually none of these drivers can be tested in experimental conditions.

I was unable to rate the comprehensiveness of the framework for three reasons. Firstly, and as already mentioned, I do not think that the proposed framework can realistically identify the possible “causes” of

reporting bias – let alone all the possible causes. Secondly, the authors constructed the typology based on a “systematic review”, which I haven’t seen. Systematic reviews of narrowly defined questions are notoriously hard to conduct; I don’t think that a review of such a broad question can be truly systematic. Thirdly, I don’t have enough knowledge of the topic to suggest missing/alternative categories.

I would rate the parsimony of the framework as medium. On the one hand, I do not think it is possible to propose a parsimonious framework when the scope of the theory is so broad. On the other hand, Table 1 is much more parsimonious than the vast literature it draws on.

I would rate the scope of the framework as fairly high. The authors acknowledge that the determinants of the model “are mostly based on research in the biomedical area”. However, I would argue that the 12 categories listed in the framework are also relevant to social research, which is the type of research that I do myself.

In conclusion, I am grateful for the authors’ contribution to the literature and do believe that the proposed framework could help the research community address the problem of reporting bias. However, I have some concerns regarding (i) the strength of the evidence used by the authors to make causal claims; and (ii) the testability of the framework.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Meta-research; research integrity; systematic reviews; social sciences; science policy; education; experimental and quasi-experimental methods.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 26 Jun 2019

Jenny van der Steen, Radboud University Nijmegen Medical Centre, Geert Grooteplein Noord 21, The Netherlands

1. I was pleased to review this manuscript. I assessed the proposed theoretical framework based on the following criteria: (1) utility, (2) comprehensiveness, (3) parsimony, (4) testability, (5) heuristic, and (6) scope. Ironically, these criteria are not part of an established theoretical framework; they only reflect a brief review of the literature on the subject.

Response: Thank you. We appreciate the reviewing of our work against a series of relevant criteria.

2. I would rate the utility of the proposed framework as high. As far as I know, this is one of the first attempts to synthesize the literature on the factors driving reporting bias. Additionally, and as mentioned by the authors, understanding these factors is essential to the development of mitigating measures.

Response: Thank you, indeed we aimed at identifying candidate determinants of reporting bias and interrelationships with a view to develop interventions in a more systematic way, and to allow for testing to improve the theoretical framework.

3. I would rate the heuristic (i.e. evidence base) of the framework as low. Granted, there is now a large body of literature on the prevalence of reporting bias and on the possible factors driving it. However, virtually none of these studies are experimental, so all the relationships found between the occurrence of reporting bias and the 12 categories are at best correlational – not causal. This is, in my view, the most fundamental flaw of the framework. The good news is, it could be easily addressed by changing the title of the manuscript from “Causes of reporting bias” to “Drivers of reporting bias”.

Response: We fully agree that the evidence base for the theoretical framework is still low. We found only 1 randomized trial in the 64 articles we analysed in the review. Numerous studies have been performed, and also interventions such as training and pre-registration have been implemented but very few interventions have been evaluated thoroughly. This is why we speak about putative determinants and plausible mechanisms. We are not sure whether “drivers” is essentially different from “causes.” Heuristic is probably in how we could learn about causes. To this end, it is important to consider the subtitle which is “a theoretical framework.” This can be developed also on the basis of a modest evidence base, because the aim is to test the model’s assumptions and adapt the model if necessary. This might progress the field at a faster pace than continuing without any or only an implicit theoretical framework. At least, we believe that considering the current evidence base, developing a theoretical framework is timely. We do recognize the small evidence base of the framework.

To the last paragraph of the Discussion, we added a sentence, to start the paragraph with explicitly recognizing the small evidence base: “Currently, the evidence for the theoretical framework is limited.”

4. Related to the above-mentioned point, I would rate the testability of the framework as low. The only testable driver of reporting bias is “financial conflict of interests” (FCOI). Most other drivers (prejudice, relationship and collaboration issues, doubts about reporting being worth the effort, etc.) would be hard to test empirically. Virtually none of these drivers can be tested in experimental

conditions.

Response: We agree that testing the model will require thoughtful consideration of putative determinants and how these could be manipulated. Models can be helpful even if not testable in a traditional way and many theories have been developed and maintained this way. Further, we need to consider how other than experimental designs can facilitate a better understanding of causes (importantly, more in-depth and thorough qualitative research which has been largely neglected in this area). The model itself shows that the issue is multifactorial which means often multicomponent interventions are required which complicates identifying contributions of single components. Creative experimental designs such as manipulating realistic scenarios withholding or adding information in different conditions and choice experiments may offer alternative ways of researching researchers in situations that are otherwise standardized. We agree that testing the model will require huge efforts, but we do not believe it is impossible.

Perhaps the focus of the Discussion was too much on experimental work. The model may also improve the quality of design of observational studies. For example, when, to assess associations with a specific determinants, the model inspires to measure important possible confounding factors.

We added to the Discussion: “The model may also assist in assessing potential confounding factors in observational work in which associations between a specific determinant and reporting bias is assessed.”

5. I was unable to rate the comprehensiveness of the framework for three reasons. Firstly, and as already mentioned, I do not think that the proposed framework can realistically identify the possible “causes” of reporting bias – let alone all the possible causes. Secondly, the authors constructed the typology based on a “systematic review”, which I haven’t seen. Systematic reviews of narrowly defined questions are notoriously hard to conduct; I don’t think that a review of such a broad question can be truly systematic. Thirdly, I don’t have enough knowledge of the topic to suggest missing/alternative categories.

Response: Our systematic review was published in [PLOS One](#) (2018). We apologize that we had referred to it only in the section explaining the model and not yet in the Background section. The search was systematic, but we analysed only a random sample of the literature because indeed the broad question did not allow to analyze all relevant literature. However, we assessed saturation of the qualitative content analyses of determinants abstracted from the literature in two ways, prospectively and retrospectively, both of which indicated that finding additional categories of determinants through analyzing more articles was unlikely. We were interested in the main categories which cover a number of specific determinants which may also differ somewhat for different disciplines. We commented on possible refinement needed for specific disciplinary fields in the Discussion section.

We added explicit reference to the review, inserting it after “We recently developed a taxonomy of putative determinants of selective reporting abstracted from the literature” ([van der Steen et al., 2018](#)). Next, we clarified the methods by several additions to the paragraph of “A framework of possible causal pathways to reporting bias.”

6. I would rate the parsimony of the framework as medium. On the one hand, I do not think it is possible to propose a parsimonious framework when the scope of the theory is so broad. On the other hand, Table 1 is much more parsimonious than the vast literature it draws on.

Response: Thank you. In the qualitative content analyses of the literature we aimed at grouping together related determinants.

7. I would rate the scope of the framework as fairly high. The authors acknowledge that the determinants of the model “are mostly based on research in the biomedical area”. However, I would argue that the 12 categories listed in the framework are also relevant to social research, which is the type of research that I do myself.

Response: Thank you. The literature we analysed predominantly concerned clinical medicine and biomedicine, but the humanities represented 11% of articles. The model also resembles the Desire-Belief-Opportunity” (DBO) model from sociology.

8. In conclusion, I am grateful for the authors’ contribution to the literature and do believe that the proposed framework could help the research community address the problem of reporting bias. However, I have some concerns regarding (i) the strength of the evidence used by the authors to make causal claims; and (ii) the testability of the framework.

Response: We hope that our theory about causality which clearly cannot reach farther than the evidence base and literature so far, inspires researchers to develop smart designs to discover causal mechanisms and to propose adaptation of the theory if needed. Thank you for this interesting and thoughtful input.

Other changes.

We edited the text at several places, shortening it where possible.

In the Discussion, we shortened the example of an intervention with regard to the means for comparing protocols and publication. Software to compare protocol and publication was not necessary in the NIHR Journal Library system and we added reference to Wright et al. (2018) for this.

Competing Interests: No competing interests

Reviewer Report 11 April 2019

<https://doi.org/10.5256/f1000research.20029.r46634>

© 2019 Bazdaric K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ksenija Bazdaric

Department of Medical Informatics, Faculty of Medicine, University of Rijeka, Rijeka, Croatia

I was happy to review a manuscript about a theoretical framework in the field of reporting bias. I think the authors have proposed an interesting perspective but my major remark is that they try to explain human behaviour with an epidemiological model for which I don't find a body of evidence in the literature that

could convince me.

Comments:

1. Background: In clinical research, registration of trials prior to data collection is used to prevent selective reporting with some success – please delete “some success” because it is further explained.
2. **"A framework of possible causal pathways to reporting bias - Motivations and means.**
Along these lines, we hypothesize that the combination of two of the most common categories in our review (van der Steen et al., 2018) — i.e., focusing on preferred findings and employing a poor or flexible study design, suffices to cause bias through selective reporting..." – how do you then comment on the replication crisis in psychology and the experiments that were replicated in the same laboratories? They were motivated to replicate and for sure were not sloppy. Do we have a poor designed field here or are there other factors? I would like a more detailed explanation^{1,2,3}.
3. A theoretical framework for reporting bias. Rothman's theoretical model – is there any evidence in practice for this model in relation to human behaviour?
Figure 1 and the model: It is an interesting figure, but the same could be explained by some other general theories, for example 'Theory of planned behaviour' (of course evidence cannot be confirmed as we have a replication crisis in psychology). I don't believe human behaviour can be explained with an epidemiological model although it is very nice. Also, the model itself does not have a word about ethical climate and other possible external factors. Why did you exclude them? Do you consider them stable in all environments?
There is some evidence that peers can predict replication²? How do you comment? Could you include some external factors in your model? Like environment, ethical climate, etc...
4. The sentence: "At the root of reporting bias may thus lay a basic human attitude, the very natural tendency to make public our successes" – this is not clear at all. At a root of everything probably lies personality and attitude, but I don't understand the meaning of the sentence here.
5. Obviously, researchers and editors are key stakeholders because commonly they decide what is actually being reported and what is not. – I would say that sometimes we cannot report everything because we have only 3000-4000 words (here is the role of editors).
6. After a series of rejections researchers may doubt whether reporting is worth the effort under the pressure of lack of resources such as time. I advise you to read a case study from Helene Speyer - The value of a “failed” trial⁴.

References

1. Baker M: 1,500 scientists lift the lid on reproducibility. *Nature News*. 2016. [Reference Source](#)
2. Camerer C, Dreber A, Holzmeister F, Ho T, Huber J, Johannesson M, Kirchler M, Nave G, Nosek B, Pfeiffer T, Altmejd A, Buttrick N, Chan T, Chen Y, Forsell E, Gampa A, Heikensten E, Hummer L, Imai T, Isaksson S, Manfredi D, Rose J, Wagenmakers E, Wu H: Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*. 2018; **2** (9): 637-644 [Publisher Full Text](#)
3. Open Science Collaboration: PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science*. 2015; **349** (6251): aac4716 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Speyer H: Discovering the value of a “failed” trial. *European Association of Science Editors*. 2018; **44** (4): 80-82

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: research ethics, psychology, medical informatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 26 Jun 2019

Jenny van der Steen, Radboud University Nijmegen Medical Centre, Geert Grooteplein Noord 21, The Netherlands

I was happy to review a manuscript about a theoretical framework in the field of reporting bias. I think the authors have proposed an interesting perspective but my major remark is that they try to explain human behaviour with an epidemiological model for which I don't find a body of evidence in the literature that could convince me.

Reponse: Thank you so much for reviewing our manuscript and considering it from the perspective of research in humanities. Thank you also for your interesting viewpoint regarding types of models, which is helpful to clarify the background of our model.

We respectfully disagree that qualifications such as *epidemiological model*, or *behavioural model* apply well to our model. The model has been developed from determinants of selective reporting mostly in the medical and biomedical literature, but we also retrieved determinants from, e.g., the humanities (11% of articles, van der Steen *et al.*, PLOS One 2018). We analyzed the determinants inductively using qualitative content analyses. This means we did not impose any previous model when we identified and categorized the determinants. Texts from the original articles inspired us to consider how determinants might work together, also distinguishing between clusters of related determinants. In a next step, we indeed recognized and used epidemiological terms to label relationships between categories of determinants in terms of effect modification and interaction. Text in the original articles also inspired us to consider whether some categories of determinants could be necessary causes, which is also terminology used in epidemiology.

However, we could just not use the terms and arrive at the same model inspired by the literature about how categories of determinants could work together. We feel that using certain terminology

does not suffice to qualify the model in such a way that it limits application to fields that do not use that terminology. The essence is our inductive way of forming categories and in part, also the thinking of relationships has been taken from the literature using an open approach about how categories may work together.

Moreover, the outcome is reporting bias in the literature which is very broad. It cannot be brought about by the behavior of a single individual. Therefore, we do not claim we can predict individual behavior. In contrast, the core of our model $\frac{3}{4}$ the necessary causes that indeed refer the most to behavior of individuals $\frac{3}{4}$ probably resembles the most a broad model from sociology ("Desire-Belief-Opportunity," DBO), which we found by coincidence after having conceived our model. Finally, the cluster of Pressures from science and society comprises four societal determinants of reporting bias and we visualized it can impact behavior of individuals directly but also indirectly.

In all, we believe there is little reason to qualify the model as belonging to a single particular discipline imposing disciplinary boundaries around its application. We rather hope that our model will cross such boundaries and help researchers from different disciplines and perhaps also policy makers to work together to address reporting bias.

In the Background paragraph of "Basis for a theoretical causal framework: hypothesized determinants of selective reporting and their interrelationships" we added that we had used an inductive approach to develop the categories. Further, we connected the paragraph with the next paragraph ("A framework of possible causal pathways to reporting bias") more clearly by starting it with "Based on what we found in the literature." Next, we reversed the order in the sentence about how we applied Rothman's classification of causes, because indeed we first agreed on the relationships and we used Rothman's labels to express the relationships in terms of his nomenclature. The sentence that started with "Inspired by Rothman's (1976) framework of necessary, sufficient and component causes," now starts with "Through multiple discussions, we inductively derived Figure 1." In the following sentence, we also clarified that we applied terms from epidemiology to what we found in an inductive way, by starting it with: "Applying more epidemiological terms to the generic model we developed."

We reflect on the approach and background of the authors by adding to the same paragraph that the discussions were "among the team with experience in both qualitative and quantitative research." Finally, to the first sentence of the Discussion, we added that our theory is "broad" and that we combined categories into clusters "inductively", "using existing epidemiologic terminology to label relationships."

Comments:

1. Background: In clinical research, registration of trials prior to data collection is used to prevent selective reporting with some success – please delete "some success" because it is further explained.

Response: We deleted as suggested "some success."

2. "A framework of possible causal pathways to reporting bias - Motivations and means.

Along these lines, we hypothesize that the combination of two of the most common categories in our review (van der Steen *et al.*, 2018) — i.e., focusing on preferred findings and employing a poor or flexible study design, suffices to cause bias through selective reporting..." – how do you then comment on the replication crisis in psychology and the experiments that were replicated in the same laboratories? They were motivated to replicate and for sure were not sloppy. Do we have a

poor designed field here or are there other factors? I would like a more detailed explanation^{1,2,3}.

Response: Thank you for your comment. The interesting Nature survey on non-reproducibility that you cited (Baker, 2016)¹ indicates that researchers believe selective reporting is a major contributor to research often not being replicable, and also a number of factors which in our model are determinants of reporting bias (e.g., pressure to publish and poor experimental design). Of course the Nature survey did not aim at offering a model or mechanisms and therefore outcomes and determinants are not being distinguished. Despite probably sharing determinants, the survey also includes “methods, code unavailable” as a reason for non-reproducibility which underlines that reporting bias is not the only reason that research is not reproducible.

Further, perhaps it was not clear enough that the model does not assert that if means and motivations are there, these are sufficient causes but this will *not necessarily* result in reporting bias. The model includes Conflicts and balancing of interest and Pressures from science and society which can modify effects on reporting bias. Further, not included in our model are interventions. Interventions may alter motivations or change the flexibility of designs and therefore diminish effects on reporting bias. We hope the model will inspire development of targeted interventions.

After “We view both clusters A and B as necessary causes, that is, they are both part of any sufficient cause of reporting bias” we added: “This does not mean that reporting bias will always be the result of presence of A and B because effects can be mitigated by interventions and modified by component causes.” Further, we added in this paragraph the term “may” in: “Together they may form a sufficient cause for reporting bias.”

The model does not specify what kind of motivations researchers may have, what would be the most interesting outcome in a particular case (e.g., in our review we found motivations to try to find the same as other research early in development of a field, but motivations later on changed to find deviations). Also we do not provide criteria for how much or how little flexibility in design would be optimal or qualify particular research as sloppy.

We further extended the paragraph with a citation from our review about motivation to report news and how the contents of the news may vary over time, in addition to citing it later on referring to development of a field: “Success can be defined in different, or even opposite ways as suggested by Rosenthal and Rubin cited by Preston et al. (2004) which was part of our review: “early in the history of a research domain results in either direction are important news but that later, when the preponderance of evidence has supported one direction, significant reversals are often more important news than further replications.” This example also illustrates the intertwining of individual motivations and societal factors.

Flexibility of design may actually be an advantage in some cases as long as the reporting about the methods is transparent, and some designs are flexible by nature. Such specifications would require empirical research.

To the last paragraph of the Discussion, we added: “Further empirical research is needed to specify, for example, what the optimal level of flexibility for a particular field and study design would be.”

3. A theoretical framework for reporting bias. Rothman’s theoretical model – is there any evidence in practice for this model in relation to human behaviour?

Figure 1 and the model: It is an interesting figure, but the same could be explained by some other general theories, for example ‘Theory of planned behaviour’ (of course evidence cannot be confirmed as we have a replication crisis in psychology). I don’t believe human behaviour can be explained with an epidemiological model although it is very nice. Also, the model itself does not have a word about ethical climate and other possible external factors. Why did you exclude them?

Do you consider them stable in all environments?

There is some evidence that peers can predict replication²? How do you comment? Could you include some external factors in your model? Like environment, ethical climate, etc...

Response: As explained above in our response to the general comment, Rothman's terms were applied later on, when the categories of determinants were already retrieved in an inductive manner from literature about determinants of selective reporting. We believe that the theory of planned behavior is not very suitable to explain reporting bias as a phenomenon based on behavior of multiple individuals, teams, institutes, peer groups, or society. How would it incorporate the categories we found in the literature, such as "Unfavourable geographical or regulatory environment" (Table 1)? Further, various other theories may apply to parts of our model, for example, from criminology, theories about strain and goal displacement.

Reviewing our dataset with 497 determinants, we find, for example, one of the determinants from the literature classified under this heading was "Increased accountability from medical journals regarding ethics and conflicts of interest" (Dieterich *et al.*, 2013; Supplemental file 2 to the review in PLOS One). It relates to the ethical climate in the publishing enterprise. Of note, at this point, we would not revise categories that are not being supported by the content analysis of the literature, but we can clarify the existing categories by providing more examples.

To further clarify the category of "Unfavourable geographical or regulatory environment" in Table 1, we added the example of "ethics in publishing enterprise"

Replication studies can help define the scope of problems with replicability which may be due to reporting bias only in part. Replication studies can also affect the four clusters A-D. It can impact Motivations (e.g., researchers becoming more interested in study results that are likely replicated or researchers conducting replication studies being more open to, or working towards in null results), Means (e.g. rise of specific journals that support publishing replicated studies), Conflicts and balancing of interests (e.g. earmarked resources for replication studies becoming available), and Pressures from science and society (e.g. less creative and innovative but rigorous research becoming more salonfähig). We added this example of replication studies impacting clusters A-D to the Discussion.

4. The sentence: "At the root of reporting bias may thus lay a basic human attitude, the very natural tendency to make public our successes" – this is not clear at all. At a root of everything probably lies personality and attitude, but I don't understand the meaning of the sentence here.

Response: we cited from an article included in the review in PLOS One because those articles provided the data we analysed to develop the determinant categories. Perhaps the reasoning was not clear enough.

Although we would not repeat the methods of the review in full, we added explicit reference to the review earlier in the text, inserting it in the second paragraph of the Background, after "We recently developed a taxonomy of putative determinants of selective reporting abstracted from the literature" (van der Steen et al., 2018).

Further, the authors of the statement find a general attitude to show success relevant for reporting bias. They do not claim to explain all human behavior. We feel the emphasis about showing success is a key issue. *We deleted "a basic human attitude" to avoid confusion about the key point.*

5. Obviously, researchers and editors are key stakeholders because commonly they decide what is actually being reported and what is not. – I would say that sometimes we cannot report everything because we have only 3000-4000 words (here is the role of editors).

Response: We agree. We found this in several articles. Table 1 therefore shows that one of twelve categories of determinants was “Limitations in reporting and editorial practices” that was explained as “Constraints and barriers to the practice of reporting relevant detail” with included among the examples “Journal space restrictions” Further, academic publication system hurdles is a separate category at the level of cluster D, science and society and it also includes editors as main actors.

6. After a series of rejections researchers may doubt whether reporting is worth the effort under the pressure of lack of resources such as time. I advise you to read a case study from Helene Speyer - The value of a “failed” trial⁴.

Response: There is a point where efforts to try to publish work are not in balance with the perceived or witnessed chances of it actually being published. This was what we found in the literature as one of the determinants of selective reporting.

To clarify the basis of the statement, We started the second section of “A framework of possible causal pathways to reporting bias,” with: “In the review, we found that..”

Obviously, trials that fail because they do not confirm the hypothesis are as valuable as trials with expected results or results hoped for. We appreciate the article of Speyer, her personal account of initial doubts and efforts to publish the results of a strong trial that in fact delivered strong evidence that the intervention was ineffective. It calls for editors and reviewers to reflect on their own possible biases and focus on the quality of the work rather than judging the quality of the work in light of their perspectives on the results and how the results may or may not differ from previous work.

We added to the Discussion: The model also shows that for example editors may influence the outcome in multiple ways; first, directly via Means (the category Limitations in reporting and editorial practices). Second, editors as a collective affect the rejection rates through Academic publication system hurdles (the category of Pressures from science and society), but also the extent to which authors find efforts to publish worthwhile (category of cluster Conflicts and balancing of interests). The latter is illustrated by the personal account of Speyer (2018).

Other changes.

We edited the text at several places, shortening it where possible.

In the Discussion, we shortened the example of an intervention with regard to the means for comparing protocols and publication. Software to compare protocol and publication was not necessary in the NIHR Journal Library system and we added reference to Wright et al. (2018) for this.

Competing Interests: No competing interest

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research